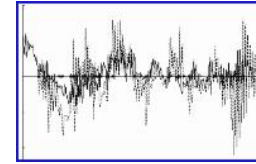




UNIVERSITÄT  
BAYREUTH

Lehrstuhl  
Mathematik VII



UNIVERSITÄT  
BAYREUTH

Mathematik VII  
Uni Bayreuth

# S4 Classes for Distributions

a software presentation  
at the Karlsruhe Stochastic Days 2004

Universität Karlsruhe

March 23rd/24th 2004

*Peter Ruckdeschel*  
*Matthias Kohl*

E-mail: [peter.ruckdeschel@uni-bayreuth.de](mailto:peter.ruckdeschel@uni-bayreuth.de)  
[matthias.kohl@uni-bayreuth.de](mailto:matthias.kohl@uni-bayreuth.de)

S4-Classes  
for Distributions

03-23/24-2004



# Outline



UNIVERSITÄT  
BAYREUTH

Mathematik VII  
Uni Bayreuth

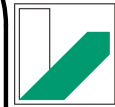
*Peter Ruckdeschel*  
*Matthias Kohl*

**S4-Classes**  
**for Distributions**

**03-23/24-2004**



Outline .....	2
1 Framework : R .....	4
1.1 What is R? — <a href="http://www.r-project.org/">http://www.r-project.org/</a> .....	4
1.2 Availability .....	5
1.3 Object Orientation .....	5
2 Package “distr” .....	6
2.1 Motivation .....	6
2.2 Designing Principles .....	8
3 Organization in classes .....	9
4 Possible Operations .....	11
5 Availability .....	12
References .....	13



UNIVERSITÄT  
BAYREUTH

Mathematik VII  
Uni Bayreuth

*Peter Ruckdeschel*  
*Matthias Kohl*

**S4-Classes**  
**for Distributions**

**03-23/24-2004**



# 1 Framework : R



## 1.1 What is R? —

<http://www.r-project.org/>

- R is a language / environment for statistical computing and graphics
- GNU project  $\approx$  S (developed by John Chambers and colleagues)
- R — “a different implementation” of S
- wide variety of statistical and graphical techniques
- R is highly extensible
- **huge** developer community





## 1.2 Availability

- R is Free Software under GNU GPL
- compiled versions for virtually any OS
- distributed through CRAN —  
<http://cran.r-project.org/mirrors.html>
- downloadable there: vast number of add-on packages

## 1.3 Object Orientation

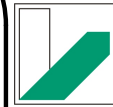
- from version 1.7.0 on, R realizes the S4-class concept
- see: J. Chambers(1998): *Programming with Data*. Springer



## 2 Package “distr”

### 2.1 Motivation

- I have an algorithm / program which works similarly for any distribution
- How to pass a distribution as an argument?
- Construction up to now:
  - a lot of distributions implemented to R — Gaussian, Poisson, Exponential, Gamma, etc.
  - for each:
    - \* cdf [ $\hat{=}$ p]
    - \* density / probability function [ $\hat{=}$ d]
    - \* quantile function [ $\hat{=}$ q]
    - \* function to simulate r.v.'s [ $\hat{=}$ r]
  - Naming convention: <prefix><Name>





- e.g. to get the median of a general distribution:

```
mymedian ← function(vtlg , ... )  
{ eval(parse(text=  
      paste("x←q" , vtlg ,  
            "(1/2 ,...)" , sep="" )))  
  return(x)}
```

- better: having a “variable type” distribution and functions  $p$ ,  $d$ ,  $q$ ,  $r$  defined for this type
- then:  $q(x)$  returns the quantile function  $\rightsquigarrow$

```
median ←function(X){q(X)(0.5)}
```

$\implies$  Development of this concept in a new package “distr”





UNIVERSITÄT  
BAYREUTH

Mathematik VII  
Uni Bayreuth

*Peter Ruckdeschel*  
*Matthias Kohl*

**S4-Classes**  
**for Distributions**

**03-23/24-2004**

## 2.2 Designing Principles

- open, easily extensible by any volunteer in the R community
- preserve naming and notation from R-base
- suggestive notation for automatically generated methods `r`, `d`, `p`, and `q`





# 3 Organization in classes

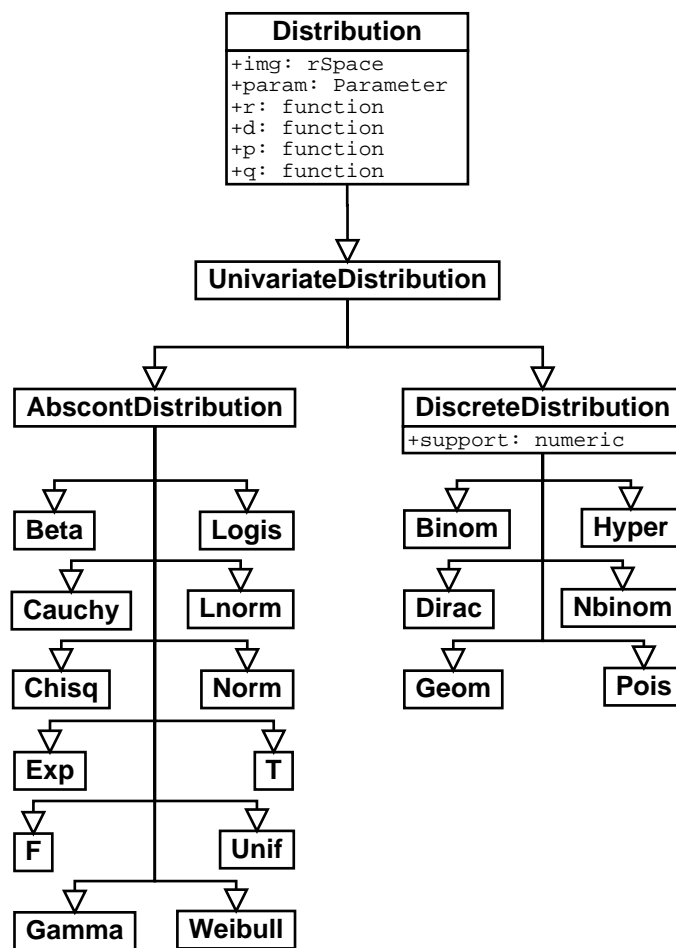
**Distribution classes:** Implementation of the concept of random variables / distributions in R. Mother class is the virtual class `Distribution` which (up to now) has the subclass `UnivariateDistribution`. The class of univariate distributions is divided up in absolutely continuous distributions with class `AbscontDistribution` and discrete distributions with class `DiscreteDistribution`.

**Parameter classes:** Classes for the parameters of distributions.

**Simulation classes:** Gather all information about a simulation in one object of class `DataClass`; has subclasses `Simulation` and `Contsimulation` (simulation of contaminated data).

**Evaluation class:** Evaluation of a new procedure (e.g. an estimator) on large set of simulation and storing of all relevant information in an object of class `Evaluation`.





slot param is filled with an object  
of the corresponding Parameter class

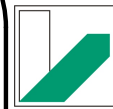


## 4 Possible Operations

**Affine linear transformation:** Expressions like  $Y \leftarrow (3 * X + 5) / 4$  for an object  $X$  of class `AbscontDistribution` or `DiscreteDistribution` giving an object  $Y$  of class `AbscontDistribution` or `DiscreteDistribution`.

**Unary mathematical operations** included in the group `math` are available so that expressions like  $Y \leftarrow \exp(\sin((3 * X + 5) / 4))$  are admitted.

**Convolution** of probability distributions is done via the exact formulas where available, respectively, via explicit calculations. For absolutely continuous summands we implemented an accurate convolution method by means of FFT (c.f. Kohl et al. (04)).





## 5 Availability

- Homepage on  
[http://www.uni-bayreuth.de/  
/departments/math/org/mathe7/DISTR](http://www.uni-bayreuth.de/departments/math/org/mathe7/DISTR)
- on CRAN soon ... —  
<http://cran.r-project.org/mirrors.html>
- once again: Homepage for R: <http://www.r-project.org/>



# References



UNIVERSITÄT  
BAYREUTH

Mathematik VII  
Uni Bayreuth

*Peter Ruckdeschel*  
*Matthias Kohl*

**S4-Classes**  
**for Distributions**

**03-23/24-2004**





UNIVERSITÄT  
BAYREUTH

Mathematik VII  
Uni Bayreuth

*Peter Ruckdeschel*  
*Matthias Kohl*

**S4-Classes**  
**for Distributions**

**03-23/24-2004**

## References

Chambers J.M. (1998): *Programming with data. A guide to the S language.* Springer.

<http://cm.bell-labs.com/stat/Sbook/index.html>. 5

Kohl M., Ruckdeschel P. and Stabla T. (2004): Computation of the convolution of arbitrary probabilities by means of FFT.  
unpublished manuscript. 5

Ruckdeschel P., Kohl M., Stabla T. and Camphausen F. (2004): *S4 Classes for Distributions.*

<http://www.uni-bayreuth.de/departments/math/org/mathe7/DISTR>. 5

